The Plymouth Student Scientist - Volume 02 - 2009 The Plymouth Student Scientist - Volume 2, No. 2 - 2009

2009

Can children use similarity between actions to learn grammar?

Anning, F.

Anning, F. (2009) 'Can children use similarity between actions to learn grammar?', The Plymouth Student Scientist, p. 80-105.
http://hdl.handle.net/10026.1/13872

The Plymouth Student Scientist
University of Plymouth

*The Plymouth Student Scientist, 2009, **2**, (2), 80-105*

Can children use similarity between actions to learn grammar?

Fiona Anning

Project Advisor: Kirsten Abbot-Smith, School of Psychology, University of Plymouth, Drake Circus, Plymouth, PL4 8AA

Abstract

The present study was a replication and extension of Casenhiser and Goldberg (2005). Children aged 5;0 to 7;6 were assigned to one of two training conditions, either they were only taught novel verbs in a novel construction paired with appearance scenes or they were additionally trained on novel verbs in the transitive construction. All participants were assessed on both word orders with new novel verbs using a forced-choice comprehension test. The children in both conditions performed significantly worse on the appearance items than the causative trials, which fails to replicate the original study. The translation post-test responses suggest that poor performance on the appearance test trials was due to the children's inability to understand the meaning of the individual verbs.

Contents

Statement of ethical compliance

The data collected for this study was obtained in an ethically sound manner. The participants were children aged between 5;0 and 7;6 and therefore the parents were asked for informed consent before their child could take part in the study. The consent form was emailed to the parents along with an information sheet which explained the purpose of the research, how long the study would take and that the children's responses would be video recorded.

Confidentiality was maintained during this research as the tape recordings were only seen by the five collaborating researchers and their supervisor. These tapes were kept in a locked room and will be destroyed after five years. The children's names were only used to assign each participant to a test order, after this point the children were only referred to by these test order numbers. The children's names and the individual scores were not included in the main report since the findings were based on the group data.

The parents were made aware of the right to withdraw their child from the research on four occasions during the study. The first opportunity was when the parent was contacted to confirm the appointment; they were reminded again when they were sent the consent form, as well as when they arrived at the Babylab before the study began. Finally when they were debriefed they were made aware that they could withdraw their child's data at any point and have it destroyed. The experiments were discontinued if the child no longer wished to respond to the researcher's questions.

All the parents were given a written debrief which explained what the researchers were investigating, this included contact numbers for the researchers and supervisor in case the parents had any further questions. The children were thanked for the help, told they did really well and given a young scientist certificate to keep. The research did not involve any deception or physical/psychological harm of the children.

The data used in this report was collected, coded (except for the post – test responses) and transcribed by the five collaborating researchers. Fiona Anning conducted the experiments for participant order numbers 3, 7, 10, 11, 12 and 18 in the replication condition and order numbers 1, 4, 9, 15 and 17 in the additional transitive condition. This researcher coded the data for the replication test orders 3, 7, 10, 12 and 18 and for the additional transitive orders 9, 15 and 17. All other test orders were investigated/coded by the other four researchers working on this project who were Sophia Starr, Alex Stimpson, Leanne Goldie and Charlotte Coles. Each researcher transcribed the post – test data for their assigned coding orders, but Fiona Anning then independently coded all of these responses.

Acknowledgements

The researcher would like to thank the supervisor Kirsten Abbot-Smith for her continual support and guidance throughout this research project. Appreciation is also given to Anthony Mee who created the computer program used to run the experiment and to the University of Plymouth's School of Psychology for the use of their Babylab database. A final thanks goes to all the children that participated in the study.

Introduction

To learn language children must be able to not only acquire word forms and grammar, but also know how they map on to the overall meaning of the construction i.e. linking syntax to semantics (for example Casenhiser & Goldberg, 2005; Fernandes, Markus, Nubila & Vouloumanos, 2006; Goldberg, 2006). There are regularities in the relationships between forms and meanings which allow children to learn these mappings before they reach their second birthday. For example Gertner, Fisher and Eisengart (2006) found that infants at 21 months could use the word order of transitive sentences to determine which of two videos matched the sentence they heard. Tomasello (2003) and Gertner et al (2006) argued that the correlations across these mappings are a major source of productivity in children (as shown by overgeneralisation errors), which in turn suggests that they have developed an abstract knowledge of a particular construction.

The two most commonly used grammatical features which determine the meaning of a construction or more explicitly 'who did what to whom' are case marking and word order (Akhtar, 1999). In English, semantic roles are marked by word order alone, so that in canonical transitive (SVO) constructions the syntactic role of subject corresponds to the semantic role of the agent and the object maps to the patient/theme (Tomasello, 2003). But, how children learn these associations is still debated.

There are two main theories for how infants come to use grammatical rules such as case marking and word order to establish the meaning of a sentence/clause and it is from these perspectives that much of the research in this field has been conducted. The first of these approaches is linguistic nativism which assumes that children are born with an innate universal grammar (UG) that allows them to learn any natural human language (Chomsky, 1965). The idea of a UG developed from two observations; the first of these is that there is poverty of the stimulus or more specifically that the language children hear (the input) is not of sufficient volume, too abstract and contains too many errors to fully explain children's rapid acquisition of grammar (Chomsky, 1965). The second claim is the continuity assumption, which is that children's language contains the same linguistic representations as adult-like grammar from the start (Chomsky, 1965).

However, a major drawback of this theory is the linking problem which is how do children link the input from their native language to the rules of UG? Especially as syntactic functions are not perceptually similar across all languages. In response to this issue Pinker (1984) proposed that children are also born with a set of innate linking rules that map experiential events in the environment to syntactic categories. The most recent theory regarding the nature of these rules is canonical linking, which suggests that children are born with mapping rules that associate semantic roles (such as the agent) with their most typical syntactic function (for the agent of an action this is the subject) (Van der Lely, 1994). Pinker (1984) claimed that the correlation between the meaning of a sentence and syntax facilitates children's grammatical acquisition in two ways. The first is via the process of forward linking (semantic bootstrapping) where cues about the meaning of a construction can provide a guide as to the probable syntactic structure. The second method is more complex and is known as reverse linking/syntactic bootstrapping, this is where children learn that specific types of verbs are more likely to occur in certain syntactic constructions (for example causative verbs in the transitive

construction), which can in turn provide valuable information regarding the meaning of the verb (in this case that an agent is causing something to happen to a patient).

The alternative approach to linguistic nativism is usage based/construction theories of grammatical acquisition, which focus on the communicative intentions of constructions as well as the form. These researchers claim that language learning uses domain general processes, rather than mechanisms specific to grammar and that like other cognitive/social skills the development of language requires that infants have experience with the input (Tomasello, 2003). Tomasello (2000, 2003) proposed that because children have to encounter each utterance on specific occasions of use to acquire them, that grammatical development is gradual rather than present from birth. Since there is no UG these theories do not have to account for the linking problem.

The verb island hypothesis is a well researched usage-based theory of grammar developed by Tomasello (1992, as cited Tomasello (2000)), which focuses on the item-based nature of children's initial constructions. The account claims that children's early linguistic knowledge is organised around individual verbs. Tomasello (1992, as cited Tomasello (2000)) investigated his daughter's (T) verb acquisition between 15 and 24 months and found that there was considerable variability between how often and in which constructions different verbs were used. He reported that like all children, T's verb-based constructions contained open slots into which nominals could be inserted. Tomasello (2000, 2003) argued that once children have acquired a number of exemplars of a construction they begin to group elements together into functionally similar categories, which then leads to productivity. The verb-island hypothesis emphasises the item-specific nature of children's early language learning, this implies that there must be factors which make some constructions easier to acquire than others.

It is clear that the two main perspectives disagree on the processes that underlie grammatical acquisition and to what extent factors might facilitate or inhibit the linking of forms to meaning. Linguistic nativists such as Pinker (1984) assume that linking rules are "near universal" and that therefore they do not need to be learnt, so no particular features would help or hinder the process. Usage based theorists however propose that children learn the mappings through their experience with language, so factors in the input may alter the learning of form to meaning correlations. A range of studies have been conducted in order to identify any variables which influence these mappings, if any effects are found they can be taken as support for the usage based view and decrease the need for innate mechanisms to explain children's acquisition of language.

A limited number of studies have found that the use of pronouns instead of/as well as full noun phrases in constructions can assist children's linking of form to meaning (Akhtar (1999); Childers & Tomasello (2001); Dodson & Tomasello (1998). For example Childers and Tomasello (2001) found that when children aged 2;6 heard pronouns and NP's during the training of familiar and unfamiliar verbs they produced significantly more transitive constructions at test compared to the children only trained with nouns. One explanation for this finding is that typically pronouns unlike NP's are case marked, which provides additional evidence alongside word order for children to use in their assignment of semantic roles (Dodson & Tomasello, 1998). Furthermore, using pronouns as the subject/object within a sentence might have reinforced the similarities between different exemplars of a construction, which would allow the development of more efficient form to meaning correlations (Tomasello, 2003). Dodson and Tomasello (1998) propose that children may centre their early constructions on pronouns as well as verbs. Together

these studies show that one way to aid children's acquisition of form to meaning mappings is to use pronouns and NP's in the nominal positions of a construction, as this highlights the similarities between sentences and can facilitate abstractions.

Previous studies have found that both type and token frequency of the input can play an important role in learning the links between syntax and semantics. Type frequency refers to the number of times a child hears one particular example of a category or construction (for example the regular past tense morpheme 'ed'), whereas token frequency is how often a type of construction such as the active transitive occurs in the input. For example Cameron-Faulkner, Lieven and Tomasello (2003) in their study of child directed speech found that children used the same NP's and copulas at similar levels of frequency as their mothers. One possible explanation for why frequency has an effect is that since language is a skill it requires practice and the more frequently a construction is heard in the input the earlier its form to meaning correlations are learnt (Cameron-Faulkner et al , 2003). The complexity of the input may also have a role in children's learning of syntactic semantic mappings, because if an adult's sentence is more complicated than the child's current linguistic skills its unlikely that they will be able to appreciate the meaning even if they are capable of production (Tomasello, 2003).

Research by Ambridge, Theakston, Lieven and Tomasello (2006), demonstrated that it is not just the frequency of the input that influences the learning of syntactic constructions, but the distribution of the input as well. The study reported that distributed exposure to a different example of the object-cleft construction each day for ten days or distributed learning of pairs of exemplars over five days was more beneficial to children's learning of the construction than if all ten examples had been encountered in just one session (massed exposure) despite the overall frequencies remaining the same. These results provide evidence in support of a general distributed learning hypothesis and suggest that children's learning of new form to meaning correlations can be improved by providing them with a more frequent but dispersed learning schedule.

Analyses of children's spontaneous speech have found that a single verb or a limited set of verbs may account for the majority of children's utterances and facilitate the learning of new form to meaning mappings. For example Ninio (1999, 2005) suggested that Hebrew and English children first learn "path breaking verbs" such as 'want' that explain basic relations among objects and can be used to facilitate future learning and generalisations of syntactic semantic correlations. Experimental support for this proposal that a single verb with high token frequency can help children generalise form to meaning relationships can be found in the research by Goldberg, Casenhiser and Sethuraman (2004). The researchers reported that when teaching adults' novel verbs in a novel construction associated with the meaning of appearance, their performance with new novel verbs at test was significantly better when the frequency of presentation had been skewed during training compared to those who received equal training on each verb. A possible explanation for why these studies have found that the learning of form to meaning mappings can be improved by using a single exemplar under high token frequency is because this emphasises the functional similarity between different instantiations, which can lead to the development of abstractions.

Other studies which have addressed the role of statistical regularities in the input have found that children use such patterns to aid their learning in many different areas of language. For example Gomez and Gerken (1999) reported that infants at 12 months could use the transitional probabilities of a novel grammar learnt during training to

discriminate new grammatical from ungrammatical word strings at test, as shown by longer looking times for the grammatical utterances. In a similar study looking at phonotactic regularities rather than word order, Chambers, Onishi and Fisher (2003) found that infants could rapidly distinguish legal from illegal syllables after a training phase. Tomasello (2003) on the basis of such research suggested that form to meaning correlations are also influenced by predictable relationships and that once children have learnt a sufficient number of item-based constructions they are able to group linguistic items which function similarly into abstract categories such as nouns and verbs. Each paradigmatic category also has certain case markers associated with it for example verbs have tense markers and nouns have plural inflections. So, statistical regularities in the input can also indirectly aid the acquisition of form to meaning mappings by providing an additional source of information on which to base these correlations.

Another factor which may influence children's acquisition of form to meaning associations is their use of semantic analogies. Gentner and Namy (2006) proposed that the mapping of verb-based constructions to their meaning may be particularly reliant on semantic analogy, because the content of the slots varies so much that only the relationships between semantic roles remain constant. A number of studies provide evidence for the theory that these comparisons facilitate the linking of syntax to semantics. Ramscar (2002) reported that the past tense inflection adults used with a novel verb was determined by how semantically similar it was to other verbs in the context given to participants. In their study Ambridge, Pine, Rowland and Young (2008) found that children aged 9 to 10 years and adults were more likely to accept as grammatical any overgeneralisations of the transitive construction to novel intransitive verbs if they belonged to a causative semantic verb class (based on Pinker's (1989) classification system where semantically similar verbs cluster together). However, there was far less of a facilitatory effect of verb class on grammaticality ratings for children aged 5 to 6 years; one explanation for this is that children at this age have not yet formed abstract narrow range verb classes and so can not use these to aid their judgements. This research showed that the meaning of constructions can help older children and adults to identify appropriate syntactic forms. These studies demonstrate that forming comparisons between the semantics of different constructions highlights their functional similarities, but that younger children are not yet proficient at this.

Numerous studies have been conducted to observe children's growing ability to use word order to facilitate their form to meaning mappings (Abbot-Smith, Lieven & Tomasello (2001); Akhtar (1999); Casenhiser & Goldberg (2005)). For example Akhtar (1999) showed that children aged 2;8, 3;6 and 4;4 were able to switch to the canonical transitive (SVO) word order when the experimenter produced familiar verbs in a non canonical order, however only the children aged 4;4 consistently switched to canonical word order with novel verbs. The research shows that the older children have an adult abstract representation of the transitive construction which can be generalised to novel forms, whereas the younger children only have item-based knowledge. The empirical study by Casenhiser and Goldberg (2005) found that children aged 5 to 7 years were able to learn and generalise (to new novel verbs at test) the pairing of novel forms with a novel meaning of appearance when presented in a non canonical English word order. This indicates that by school age children can rapidly acquire and extend new form to meaning correlations which are associated with a common novel word order. Together these experiments show that for children over four years the order of words in a

sentence provides valuable information about the meaning, because they have developed abstract representations of the most common constructions.

The present research is a replication and extension of the second study conducted by Casenhiser and Goldberg (2005). These experimenters looked at whether children aged between 5;0 and 7;6 trained on novel verb to meaning mappings associated with a novel word order (SOV with appearance) would perform the same as they would with the mappings of familiar verbs to their meanings in the canonical SVO construction or whether the children had just learnt to select the appearance scene in the first study. The results showed that the participants who had received training on the novel construction demonstrated an equal level of knowledge for both transitive and novel sentences at test, compared to the control children (who received no training on the novel word order) who performed significantly better on the transitive constructions.

The current research will replicate the work of Casenhiser and Goldberg (2005) by presenting children aged 5;0 to 7;6 with novel verbs associated with the novel meaning of appearance. With school aged children it is important to use novel and not familiar verbs to show that children develop a truly abstract knowledge of the novel word order, rather than their performance being aided by verb specific knowledge based on previous experience with a known verb in a variety of constructions (Akhtar, 1999).

Casenhiser and Goldberg's (2005) first experiment reported that skewed frequency of the training input had a facilitatory effect on children's learning of form to meaning mappings; as a result of this finding they only used skewed frequency in their second study. To replicate this study the current research will present children with verbs where the frequency of presentation has been skewed. During the training phase participants will be shown eight sentences paired with eight different film clips, which will contain five novel verbs, one of which will be heard four times whereas the others will only be played once. In the appearance training the video clips will show a puppet spontaneously appear into the scene in a different way for each novel verb.

The current experimenters will train half the children on only the novel appearance construction (SOV), as conducted by Casenhiser and Goldberg (2005) in the original experiment. The other children will be trained on both this novel word order and the causative transitive construction (SVO). The causative training will be alternated with the appearance training and will be made up of a further eight sentence and film clip pairings, in which verb frequency will again be skewed. In this condition the novel verbs will refer to a novel action that one puppet will cause a different toy to perform, importantly these characters will remain in the scene throughout. The actions have to be novel since by 5;0 to 7;6 the children will know familiar causative verbs such as 'push'.

Children in both conditions will be given the same forced - choice comprehension test, in which they will be instructed to point to the scene that matches the sentence being played (all of which will contain a new novel verb). Two film clips involving the same characters will be presented side by side; one clip will show a new appearance scene and the other a new causative action. Which of the two clips is correct will be determined by whether the children hear the verb in the novel word order (SOV) or in the transitive construction (SVO). In an extension of Casenhiser and Goldberg (2005) the children will then be given a translation post test in order to investigate what they interpret the novel verbs to mean in English and to explain any null results.

If the children in the replication condition are able to acquire and generalise the form to meaning mappings associated with the novel word order (i.e. SOV relates to

appearance) they will perform equally well at test on both the novel and transitive construction with new novel verbs. If they are unable to extend the meaning of appearance to new novel verbs in the non canonical order, then they are likely to perform better on the transitive test trials than on the appearance items. It is possible that being simultaneously trained on two different word orders might confuse the children in the additional transitive condition which could impair their performance on the novel (SOV) trials; their performance on the transitive construction ought to be unaffected as this should already be acquired by 5;0 – 7;6 (Akhtar, 1999).

Method

Participants

40 monolingual English speaking children aged between 5;0 and 7;6 (mean age 5;8) were recruited through the University of Plymouth Babylab database. 16 of the participants were male and 24 were female. Parents were asked to report whether their child had ever had a hearing or developmental impairment which might have affected their language development or whether their child had ever been diagnosed as having a language delay. No other biographical data was obtained.

Design

A two factor mixed design was used where the between subjects independent variable was the type of training the participants received and the within subjects variable the type of clip. The percentage of correct scores was the dependent variable. The experiment consisted of four separate phases: practice, training, pointing test and translation post-test. The practice session was the same for all participants and consisted of six pointing trials, where the children were instructed to touch one of two simultaneously presented films which corresponded to the sentence they heard. For the training phase the children were randomly allocated to one of two conditions. The first condition was a replication of Casenhiser and Goldberg's (2005) second experiment, in which participants had to attend to a series of eight film clips where a novel verb in the novel word order (SOV) was paired with a scene of appearance. In the other condition (additional transitive) the children were trained on eight scenes of causation, which were associated with a novel verb in the English transitive word order (SVO), as well as the appearance training. The children in both conditions then received the same test, in which they were presented with a set of six trials where like in the practice phase they had to point to the clip which matched what they heard; for three of the pairs the appearance scene was the target, whereas for the other three the transitive clip was the target. The children were then shown all of the clips they had been presented with during the training (eight for the replication children and 16 for the additional transitive group) and pointing test (a total of 12, as they saw six sets of pairs) without the accompanying sentences and were asked to describe what they thought was happening in the scene.

The experiment was conducted by five different researchers, but each experimenter used the same materials and a standardised set of instructions (see Appendix A). The researchers were also randomly allocated to different test orders across both the replication and additional transitive conditions, this meant that they all

tested both males and females and so any effects that multiple experimenters may have had on the results were minimised.

Procedure

Before attending the study the children's parents were sent an information sheet and consent form attached to an email (see Appendix B), which they were asked to read, sign and bring to the experiment in order to give consent for their child to partake in the experiment. Each child was tested independently, although parents and siblings could remain in the room provided that they did not attempt to influence the participant's responses in any way. The child was asked to sit directly in front of the computer screen with a video camera positioned behind them that was angled towards the screen in order to film their pointing responses. A microphone was attached to the camera to record the children's verbal responses to the translation post-test clips. These recordings could then be coded later by the researchers (for the coded raw data view the attached disk). Participants were informed by the experimenter that the computer screen was a special screen that they could touch. A computer program was developed which ran the film clips alongside their corresponding audio sentences for each stage of the study. All the audio clips were recordings of a female with a regional (Plymothian) accent, which was likely to sound familiar to the children.

The first phase of the experiment consisted of six practice sessions (always presented in the same order) in which all the children saw two clips of puppets performing similar actions presented side by side on the screen at the same time as hearing a sentence which described just one of the clips. The children were asked to touch the clip that the lady was talking about once she had finished talking and were given feedback after each response.

Once the child had completed the practice sessions the experimenter moved onto the training trials, where the child was told by the researcher that they would now hear the lady talking in an alien language. For this stage of the experiment the participants were randomly allocated to one of two between-subjects conditions. The first of these conditions was a direct replication of Casenhiser and Goldberg's (2005) second experiment and trained the children on novel verbs in the novel word order (SOV) which was associated with scenes of appearance. The participants were shown eight appearance film clips paired with an audio track that described the scene using the novel construction in the present tense and once the clip had stopped in the past tense as well; all the training trials were then repeated. The order of presentation of these verbs was counterbalanced across the children. The children heard a total of five novel verbs which were presented with a skewed frequency, so that one verb accompanied half the clips and the other four verbs were only heard once each. For each novel verb a puppet spontaneously appeared in the scene in a different way. All five appearance related novel verbs ended with the morpheme 'o', before any tense/person agreement was added (for a list of all the training sentences used in the replication condition see Appendix C).

In the second (additional transitive) condition the children were trained on both the novel word order (SOV) as described above and the causative transitive construction (SVO) (i.e. they received extended training compared to the replication condition). These children were presented with a further eight film clips of puppets performing causative transitive actions paired with audio tracks using standard English

transitive word order (for a list of the causative training sentences see Appendix D). The novel verbs in this condition referred to a novel action which one puppet caused another toy to perform, the puppets in these clips remained in the scene throughout the duration of the film, so that the meaning was purely causative and not of appearance. The transitive sentences contained five novel verbs presented with the same skewed frequency as those in the appearance tracks, but these verbs did not end with any morphological marker. The training alternated between the two types of sentences (which were both counterbalanced across the children) and the whole training session repeated so that the children in this condition saw 32 film clips in total.

The children in both conditions were then presented with the same forced-choice comprehension test made up of six trials, as used in the study by Casenhiser and Goldberg (2005). For this part of the experiment parents were asked to close their eyes so that they did not influence their child's responses. For the same reason the researcher avoided looking at the screen during this stage of the study and instead focused on the child. The children were simultaneously presented with two new film clips on the screen both involving the same puppets, but one scene showed an appearance and the other clip a causative action. Just like in the practice sessions the children only heard one sentence (only in the present tense) and were instructed by the experimenter to touch the clip that corresponded to what had been played (for a list of all test sentences see Appendix E). All children were told not to touch the screen until the lady had finished talking to ensure that they had heard the whole sentence. For three of the items the appearance clip was the target and the causative the foil, for the other three trials the opposite was the case. The order of presentation of the test clips, whether the causative or appearance clip was the target and whether this was presented on the left or right side of the screen were all counterbalanced across both groups of children to prevent any systematic biases in their responses.

Once the child had finished the six test trials their parent(s) were allowed to open their eyes and the researcher moved on to the final stage of the study: the translation post-test. The children were shown all of their training (8 for replication participants and 16 for the additional transitive condition) and test clips (12) again one at time without any audio and were asked to say what they thought the puppet(s) was/were doing in English, to establish whether they could correctly associate appearance verbs with scenes of appearance and causative verbs with causative actions. All the children received the same feedback ('good') after providing each description, regardless of whether or not their response was correct. If a child did not give an answer for a particular scene the experimenter moved on to the next clip.

After the translation post-test the study was finished and both children and parents were thanked for their participation. Parents were given a debrief (see Appendix F) which explained the purpose of the study and contained contact details should they have any further questions or wished to withdraw their child's results. £3 was given to parents to cover travel expenses and a young scientist certificate given to the child.

Coding

The tapes from each of the children were played back and coded after the experiment had finished. Each of the five experimenters coded the recordings of eight children. The practice and test trials both required participants to touch the scene which corresponded to what they had heard and were therefore coded in the same way (for the coded raw

data view the attached disk). Participants got a score of 1 if they touched the correct clip, 0.5 if the touched both and 0 if they selected the incorrect scene. N/A was recorded for any responses which could not be coded, either because the child did not respond or their head blocked the camera's view of the computer screen. For the practice trials each child received a score out of a total of six. For the test trials the children were scored separately for the appearance and causative items with a maximum of three for each sentence type, so that their performance on the two different sentence types could be compared. For the test items percentage correct scores for both sentence types were also calculated.

The supervisor Kirsten Abbot-Smith (KAS) coded 65% of the children's test trials blind to how the experimenters had coded them. There was perfect agreement between her coding and that of Sophie Starr and Fiona Anning (Cohen's Kappa = 1.0 or 'perfect agreement' for both). The agreement between KAS and Leanne Goldie, Alex Stimpson and Charlotte Coles were 91%, 83% and 70% respectively. Overall, there was 88% agreement (Cohen's Kappa = 0.79).

The verbal responses given by the children in the translation post-test were transcribed by all the researchers and then independently coded by Fiona Anning. The answers were given a score of 1 if they were correct and 0 if incorrect. The additional transitive training clips were only viewed by the children in this condition and therefore no scores were given to the children in the replication condition for these films. The classification used to determine a correct response depended on the type of clip being described. For the appearance (SOV) test and training clips a response was defined as correct if either the child said the word 'appear' or if they said a movement verb plus a preposition of change of location (except if going back in again was mentioned), all other descriptions were coded as incorrect. A response was coded as correct for the transitive test and training scenes only if the child used a verb that referred to the action carried out by the agent that directly affected the target patient. Percentage correct scores were calculated separately for each of the clip types.

Results

This section will first examine the data obtained from the practice and test trials, before analysing the children's responses to the translation post-test (for all raw data view attached disk). For both the test and post-test data, the analysis will initially be conducted on the children's overall performance across all the verbs and then on each verb individually.

Practice trials

The mean number of practice trials answered correctly was calculated separately for the replication and additional transitive conditions to identify whether both groups understood the demands of the task equally well. The results indicate that all the children performed close to ceiling across the practice items, with a mean score of 5.45 (out of 6) for the replication condition and 5.48 for the additional transitive condition (see Appendix G for the SPSS tables of the practice trial descriptive statistics). Since all the

children were deemed to have sufficient understanding of the task requirements none were excluded from progressing on to the training and test phases of the study.

Test trials

Overall performance

Table 1 displays the descriptive statistics for the test trial data for each word order under both conditions (see Appendix H for the SPSS table of the forced – choice comprehension test descriptive statistics).

Table 1. *Mean and standard deviation of percentage of correct pointing trials for both the SOV and SVO word orders in the replication and additional transitive condition (n=20).*

Word order	Condition	Mean	SD
SOV	Replication	.35	.35
	Additional transitive	.26	.24
SVO	Replication	.72	.3
	Additional transitive	.74	.24

Table 1 shows that the children had higher percentage correct scores for the transitive (SVO) sentences than for the novel (SOV) construction, in both conditions. The mean score for the SOV word order in the replication condition is slightly bigger than the mean score in the additional transitive condition. There is very little difference in the scores across the two conditions for the SVO sentences. A high level of variance is reported for all the conditions, which implies that there was variability across the children in their understanding of both the novel and the transitive construction.

A two factor mixed design ANOVA identified a significant main effect of word order on percentage correct scores, $F(1, 38) = 46.62$, $p < .001$ (see Appendix I for the SPSS ANOVA tables). This indicates that the children's percentage of correct trials on the novel construction differed from their performance on the transitive word order items. This fails to support the prediction that the children will perform as well on the novel word order (SOV) as they would on the English transitive (SVO) construction. This also fails to replicate previous research by Casenhiser and Goldberg (2005), which found that children trained on the novel word order performed equally well on the SOV and SVO test trials. The main effect of condition ($p = .574$) and the interaction ($p = .361$) were not found to be significant. This means that the children in both groups performed similarly on the test trials for each word order, regardless of whether or not they had received additional transitive training. This finding fails to support the prediction that children in the replication condition will perform better on the SOV word order than the children who received additional transitive training.

A series of ANCOVA's were conducted on the data to determine whether a covariant could explain the present findings and the study's failure to replicate the work of Casenhiser and Goldberg (2005). However, no significant effect for age, gender,

experimenter, coder or performance on the practice trials as covariates were found (all p >.05) (see Appendix J for the SPSS table of the ANCOVA's). This means that none of these uncontrolled variables are responsible for the obtained pattern of results.

It must be noted that a Shapiro-Wilk test of normality reported a significant result for each word order under both conditions (all p ≤.010), which means that the scores are not normally distributed (see Appendix K for the SPSS table of normality). The data could be transformed to decrease the risk of a type 1 error, but because the only significant result was highly significant (p <.001), this was not felt necessary by the researcher.

Since there was no main effect of condition the results for the replication and additional transitive groups were combined and a one-sample t-test conducted on each of the word orders to determine whether the children's performance differed significantly from chance. For the novel (SOV) word order the percentage of correct trials (selecting the scene of appearance) was found to be significantly less than chance, t (39), = -4.07, p <.001 (two-tailed), with a mean score of .31 (see Appendix L for the SPSS table of the one – sample t – tests). This suggests that for the majority of the trials where the appearance scene was the target the children were actually selecting the causative film instead. The children's performance on the transitive word order (SVO) was significantly greater than chance (M = .73), indicating that as expected they were already familiar with the English transitive construction by 5;0 to 7;6 and so were able to correctly point to the scene of causation, t (39), = 5.49, p <.001 (two tailed). One possibility for why the children were more likely to select the transitive than the appearance scene (regardless of which clip was the target) is that they had a bias towards the causative films.

Performance on individual verbs

Table 2 shows the descriptive statistics of the test trial data for each individual verb in both word orders (see Appendix M for the SPSS table of the descriptive statistics for the individual verbs).

Table 2. *Mean and standard deviation of percentage of correct pointing trials for each verb in both word orders.*

Word order	Verb	n	Mean	SD
SOV	Shanno	16	.16	.3
	Lemmo	23	.24	.42
	Veemo	16	.47	.5
	Pooko	24	.25	.42
	Zoopo	31	.29	.42
	Neebo	16	.67	.44
SVO	Shann	24	.9	.29
	Lem	17	.9	.2
	Veem	22	.59	.5
	Pook	15	.93	.26
	Zoop	10	.9	.32
	Neeb	23	.28	.45

The analysis of the children's overall performance found that they scored significantly greater than chance on the SVO construction, but significantly less than chance on the novel (SOV) word order. Analysis was run on each verb (combined across both conditions) for both word orders to determine whether these effects were universal across each of the individual verbs. Table 2 shows that in general the mean percentages of correct pointing trials for the verbs in the SOV construction are considerably lower than the means for the verbs in the SVO word order. However, for 'neebo' and 'neeb' the means show the opposite pattern, with the verb in the SVO construction having a mean percentage correct score less than the verb in the novel word order. Furthermore the mean scores for 'veemo' and 'veem' are less extreme than the scores of the other verbs in both construction types. The degree of variance reported for each verb is very high, which suggests that there was considerably variability across the children in their understanding of the form to meaning mappings for the different verbs.

One-sample t-tests (two-tailed) were conducted to determine whether the percentage correct scores for each verb in both constructions were significantly different from chance. For the novel (SOV) word order the researcher found that the percentage of correct pointing trials for 'shanno', 'lemmo', 'pooko' and 'zoopo' were all significantly less than chance (all $p \leq .01$) (see Appendix N for the SPSS table of the t – tests on all the individual verbs). This means that for these verbs the children were more likely to select the incorrect transitive scene than the appearance film; a finding which is in favour of the causative bias explanation. The scores for 'veemo' and 'neebo' were not shown to differ significantly from chance (both $p > .05$). These findings suggest that the children's overall performance on the SOV construction was significantly less than chance as they had a poor understanding of the form to meaning associations for the majority of the individual verbs tested in this word order.

One-sample t-tests carried out on the verbs in the transitive (SVO) construction reported that the percentage correct scores were significantly greater than chance for the verbs: 'shann', 'lem', 'pook' and 'zoop' (all $p \leq .03$). The mean percentage correct scores for 'veem' were not found to significantly differ from chance ($p = .406$). For 'neeb' the mean scores were significantly different from chance, $t(22), = -2.328$, $p = .03$, however, this was in the opposite direction to the other significant results for this word order (i.e. less than chance) as shown by the mean score in table 2. This result demonstrates that the children had acquired the structure meaning mappings for the majority of the verbs used to test the SVO construction, which meant that their overall performance on this word order was significantly greater than chance.

Translation post-test

To follow-up this finding that the difference between the children's overall performance across the two word orders was dependent on their understanding of the individual verbs used in each construction, an analysis was conducted on the verbal responses given by the children in the translation post-test. Post-test data was unavailable for one of the children in the replication condition due to experimenter error ($n = 19$, $N = 39$). The children's descriptions of the scenes involving each verb were analysed for the 12 test (six appearance and six causative) and the eight appearance training verbs. The

eight causative training verbs were removed from the analysis, because they were only viewed by the children in the additional transitive condition.

Table 3 shows the descriptive statistics for each of the three clip types being analysed under both the conditions (see Appendix O for the SPSS table of the descriptive statistics for the translation post – test).

Table 3. *Mean and standard deviations of the percentage of correct verbal responses given for each clip type in both conditions (n=19 for replication and n=20 for additional transitive).*

Type of clip	Condition	Mean	SD
SOV test	Replication	.29	.21
	Additional transitive	.38	.28
SOV training	Replication	.23	.18
	Additional transitive	.38	.22
SVO test	Replication	.79	.12
	Additional transitive	.78	.12

Table 3 shows that the children gave more correct verbal responses to the SVO test items under both conditions than they did for either the SOV training or test clips. The mean scores differ very little between the novel (SOV) training and test. For the transitive test items there was very little difference between the percentages of correct verbal responses across the two conditions. For both the SOV types of clip, the participants in the additional transitive condition performed better than those in the replication condition. The degree of variance is smaller for the SVO items than for both the SOV training or test clips, which suggests that there was less variability in the children's answers for the already familiar transitive construction items than in their responses to the novel (SOV) clips.

A two factor mixed design ANOVA identified a non significant main effect of condition ($p = .07$) and of the interaction ($p = .17$) (see Appendix P for the SPSS ANOVA tables for the translation post - test). The main effect of clip type on the percentage of correct verbal responses given by the children was significant, $F (2, 36) = 86.06$, $p < .001$. An LSD post hoc analysis was conducted on the main effect of clip type; this showed that there was a significant difference (both $p < .001$) between the SVO test clip percentage correct scores and the scores for both types of appearance clips (test and training) (see Appendix Q for the SPSS table of the LSD post – hoc analysis). No significant difference was found between the scores for the SOV test and SOV training items ($p = .56$). These findings show that the children in both conditions gave a higher percentage of correct verbal responses to the SVO clips than the novel (SOV) items and

therefore had a better understanding of the form to meaning mappings for the transitive construction. This supports the results from the test trials which found that the children correctly pointed to the SVO scenes more often than they did to the SOV films.

Once again it must be mentioned that a Shapiro-Wilk test of normality reported a significant result for four of the six clip type condition pairings (all $p < .02$), which means that the majority of the scores are not normally distributed (see Appendix R for the SPSS table of normality). Only the data from the SOV test clips in the replication condition and from the additional transitive condition during the SOV training are normally distributed (both $p > .05$). The data could be transformed to decrease the likelihood of a type 1 error, but because this is unlikely to change the pattern of results the researcher chose not to.

One-sample t-tests were conducted on the post-test data to establish whether the children's percentage of correct verbal response scores for each clip type differed significantly from chance. This also allowed for the investigation of a likely bias in the children towards interpreting the scenes as having an alternative meaning (possibly causative) to appearance on SOV trials and training. The data was combined from the replication and additional transitive groups because there was no significant main effect of condition. For the appearance (SOV) test clips the children's performance was found to be significantly worse than chance, $t (38), = -4.25$, $p < .001$ (two-tailed), with a mean score of .33 (see Appendix S for the SPSS table of one – sample t – tests). On the SOV training clips the children also scored significantly less than chance ($M = .31$), which suggests that the they had not fully acquired the mapping of the SOV construction to the meaning of appearance and were therefore providing alternative interpretations of these scenes , $t (38), = -5.61$, $p < .001$ (two-tailed). As expected the children performed significantly better than chance on the SVO test clips, $t (38), = 14.83$, $p < .001$ (two-tailed), with a mean percentage correct score of .79. This indicates that they were familiar with the association between the transitive construction and scenes of causative action and were able in the majority of cases to correctly describe the SVO clips as showing causation.

Performance on individual verbs

The analysis of the children's overall performance on the post-test suggested that the children were producing alternative interpretations to appearance for the majority of the SOV test and training clips. Analysis was run on each verb individually for the SOV test and training and for the SVO test scenes to establish whether this pattern of incorrect descriptions of the clips was universal across each of the verbs. Table 4 shows the descriptive statistics from the post-test verbal responses for every verb in the SOV test and training and SVO test (once again the SVO training verbs were excluded as only the children in the additional transitive condition viewed these films) (see Appendix T for the SPSS table of descriptive statistics for the individual verbs).

Table 4. *Mean and standard deviations of the percentage of correct verbal responses given for each individual verb for the three clip types (N=39).*

Type of clip	Verb	Mean	SD
SOV test	Shanno	.69	.47
	Lemmo	.1	.31
	Veemo	.13	.34
	Pooko	.59	.5
	Zoopo	.23	.71
	Neebo	.26	.44
SOV training	Moopo (parrot)	.05	.22
	Moopo (duck)	.1	.31
	Moopo (bear)	.05	.22
	Moopo (fish)	.1	.31
	Vako	.51	.51
	Keebo	.51	.51
	Koufo	.33	.48
	Faygo	.79	.41
SVO test	Pook	1.0	.0
	Veem	.85	.37
	Shann	.82	.39
	Neeb	.13	.34
	Lem	.97	.16
	Zoop	.95	.22

Table 4 shows that for the SOV test verbs the mean percentages of correct verbal responses (i.e. an appearance explanation) are low for 'lemmo', 'veemo', 'zoopo' and 'neebo'. The mean scores for 'shanno' and 'pooko' are considerably higher at over .5. The mean percentage of appearance descriptions given for all of the 'moopo' scenes in the SOV training phase are very low, suggesting that almost all the children failed to interpret these clips as demonstrating appearance. For 'vako' and 'keebo' the mean percentage of correct responses was approximately half. A low mean score was reported for 'koufo', but a high mean score for 'faygo'. The mean percentages of correct answers for the SVO test clips (i.e. a causative description) are all very high and a lot greater than the scores for the SOV test and training scenes, except for 'pook' which had a very low mean percentage of causative interpretations. The degree of variance is very high for the majority of the clips, indicating that there was considerably variability amongst the children as to whether or not they correctly interpreted the meaning of the verb in each scene.

Two-tailed one-sample t-tests were conducted to determine whether the mean percentages of correct responses (either appearance or causative depending on the clip type) for each individual verb were significantly different to chance. For the SOV test clips the mean percentages of appearance descriptions were significantly less than chance for 'lemmo', 'veemo', 'zoopo' and 'neebo' (all $p \leq .022$) (see Appendix U for the

SPSS table of the t – tests for the individual verbs). The score for 'pooko' did not differ significantly from chance ($p = .268$). The mean percentage of correct responses for 'shanno' was significantly greater than chance ($p = .014$). These findings show that for the majority of the verbs used in the SOV test clips the children had not associated the novel verbs with the meaning of appearance and were providing alternative interpretations (possibly causative) instead. The above findings on the children's interpretations of the appearance test clips provides mixed support as an explanation for the children's performance on the pointing trials. Their incorrect descriptions of 'leemo' and 'zoopo' may explain why the children performed significantly below chance on these verbs during the pointing trials. However, the children's poor understanding of 'neebo' does not explain why they scored above chance on this verb at test. Furthermore the participant's appearance descriptions of the verb 'shanno' were above chance, yet on the test trials their mean score was significantly less than chance.

One-sample t-tests carried out on the SOV training verbs found that the mean percentage of appearance descriptions were significantly less than chance for all four clips using the verb 'moopo' (all $p < .001$) and for 'koufo' ($p = .036$). This shows that the children had not correctly learnt the meaning of the most common training verb; this finding could be responsible for their later poor performance on the appearance test verbs. The scores for 'vako' and 'keebo' did not significantly differ from chance ($p = .875$). The mean percentage of appearance descriptions given by the children for 'faygo' was significantly greater than chance ($p < .001$). So, for all but one of the SOV training verbs the children were no better than chance at correctly providing a description of appearance, because they had developed an alternative interpretation of the clip.

For the SVO test verbs all the mean percentages of correct causative descriptions (except for 'neeb') were significantly greater than chance (all $p < .001$). This finding shows that in general the children had correctly understood these verbs to mean causation. The mean score for 'neeb' was found to be significantly less than chance ($p < .001$). The children's poor performance for 'neeb' is likely to be because the clip itself was not a clear presentation of causation. These results for the children's interpretations of the SVO test clips help to explain why for the pointing trials the children performed above chance on all the verbs except 'neeb', where again the mean score was significantly below chance.

In sum it is most likely that the children's poor performance on the test trials for the SOV construction results from their incomplete acquisition of the form to meaning mapping between the novel word order and appearance, as demonstrated by their limited ability to give correct verbal responses when asked to describe the SOV clips. Whereas their relatively good performance on the SVO pointing trials is because they are already familiar with the association between the transitive construction and the meaning of causation, seen by the high percentage of correct verbal responses the children gave for the SVO items.

Discussion

The results fail to support the experimental hypothesis that children in the additional transitive condition will perform worse on the appearance test trials than those in the replication condition, because of the possible confusion caused by being simultaneously trained on two different constructions. The forced – choice comprehension test data showed that the children in both conditions performed similarly to one another on each of the word orders, i.e. there was no significant main effect of training condition or an interaction between condition and word order. This indicates that the presentation of more than one word order during training did not impair the children's performance at test compared to children only trained on a single construction.

However, the results of the test trials did show a significant main effect of word order; more specifically that the children in both conditions had higher mean percentage correct scores for the transitive test clips than for the appearance items. Further analysis which compared the children's performance on each word order during the forced – choice comprehension test to chance found that the mean percentage of trials correct was significantly greater than chance (50%) for the SVO items but significantly less than chance for the appearance clips. This suggests that the children in both training conditions were unable to acquire and generalise the novel form to meaning mappings associated with the SOV construction and consequently often failed to point to the appearance film when it was the target clip. The children demonstrated acquisition of the transitive construction and an ability to generalise this knowledge to new novel verbs at test by correctly selecting the causative scene on transitive trials. These findings fail to support the prediction that the children trained on the novel construction would perform equally well at test on the appearance (SOV) and transitive (SVO) word orders with new novel verbs. This result also fails to replicate the previous research conducted by Casenhiser and Goldberg (2005), which found that children aged 5;0 – 7;6 trained on the novel SOV construction performed equally well with new novel verbs in both the appearance and transitive constructions in a similar forced-choice comprehension test.

There are a number of possible explanations for why the current study failed to replicate the findings of Casenhiser and Goldberg (2005). The first argument to be explored was that the children failed to understand the demands of the task. However, this explanation is unlikely for a number of reasons. All the children took part in a practice session identical to the forced – choice comprehension test in that they were asked to point to the clip that corresponded to what they had heard, except that common English rather than novel verbs were used and the films were of two different actions not one causative and one appearance scene. Overall performance on the six practice trials was close to ceiling for both conditions and no individual child scored less than four correct, this was taken to indicate that they had sufficient understanding of what was required during the task. Three pairs of clips used in the practice phase of the current study were those used by Casenhiser and Goldberg (2005) in their experiments. Further to this, statistical analysis found that the children's scores on the practice items was not a covariate of their performance on the test trials, which means that those children who performed poorly on the forced – choice comprehension test did not necessarily score lower on the practice items. It therefore seems very unlikely that a poor understanding of the task was responsible for the failure to replicate the original study.

Another possible reason for why the children in the current study failed to perform as well on the appearance items as they did on the transitive trials is that their mean age was lower than the average age of the children in Casenhiser and Goldberg's (2005) second experiment, although the range used was identical (5;0 – 7;6). In the current research the mean was 5;8, whereas in the older study the mean age of the children was 6;3, this shows that on average the present sample were seven months younger than those that participated in the Casenhiser and Goldberg (2005) study. It could be argued that this difference in mean age between the two groups may explain the null findings since grammatical development continues throughout early childhood until at least eight years (Tomasello, 2000). However, an ANCOVA conducted on the results found that age was not a covariate of the children's forced – choice comprehension test performance; more specifically that the older children did not have higher mean percentage correct scores than the younger participants. For this reason the difference between the mean ages of the two samples is probably not responsible for the current study's failure to replicate the findings of Casenhiser and Goldberg (2005).

To ensure that none of the five different researchers that conducted the experiments and coded the data had a significant effect on the children's performance, which may have caused the null findings, both the experimenter and coder codes were run as covariates. The statistical analysis demonstrated that both variables had non significant effects on the children's performance and therefore it is unlikely that the use of multiple researchers was the reason why the children in the current study often failed to select the appearance scene on novel (SOV) test trials.

Another factor which may be responsible for the children's poor performance in the present study is that they failed to understand the form to meaning mappings of the individual verbs used in the appearance test trials. There are a number of different lines of evidence in favour of this interpretation. The first indicator of the children's inability to recognise the correlation between syntax and semantics for the SOV verbs was that they selected the scene of appearance significantly less often than chance for four of the six SOV test verbs ('shanno', 'lemmo', 'pooko' and 'zoopo'). This indicates that for the majority of the appearance test verbs the children were unable to use semantic analogy to acquire the mapping between the SOV word order and the appearance scene and were instead pointing to the causative clip. Whereas when the same four verbs were used in the transitive construction the children chose the causative film significantly more frequently than chance. These findings suggest that in general the children were biased towards selecting the causative scene on every trial regardless of whether they had heard the verb in the novel or transitive construction. Evidence of a bias for transitive actions has been identified in previous research by Papafagou, Cassidy and Gleitman (2007), which found that children aged 3;7 – 5;9 and adults preferred to make action attributions over interpretations of desire or belief. Boyd, Gottschalk, Goldberg (in press) also reported a bias in adults towards novel verbs in word orders that have more in common with their native (English) constructions such as theme – before – locative than word orders that are less frequently found in English. So, it is plausible that the children in this study had falsely associated the majority of the verbs used in the novel construction as well as those in the transitive sentences to the causative clips, resulting in low mean percentage correct scores for the SOV test verbs.

Further support for the theory that the children's poor performance on the appearance test verbs resulted from their limited ability to correctly associate form to

meaning using semantic analogy for each of the individual verbs, is evident in the verbal responses they gave during the translation post – test. The post – test provided the researchers with specific information about how the children interpreted each of the test clips for both construction types. Overall the proportion of correct descriptions of each scene (appearance for the SOV clips and causality for the SVO items) was significantly greater than chance for the causative films, but significantly less than chance for the appearance items. This finding clearly indicates that the children had developed interpretations of the SOV films that were not based on appearance and had therefore failed to understand their true meaning. However, it is unclear what the semantics of these alternative interpretations were, since the responses were only coded as correct (i.e. appearance) or incorrect (any other answer) by the researchers. Future studies should aim to code the incorrect responses according to their meaning, so that the nature of the children's understanding of the clips can be identified and the possibility of a causative bias confirmed.

The analysis conducted on the responses given by the children in the post-test for the individual verbs used in both word orders, provides mixed support for the idea that because they failed to correctly understand the form to meaning mappings for the SOV verbs the children performed poorly on these items during the forced – choice comprehension test. The mean percentage of correct descriptions given for the appearance test items was significantly less than chance for four of the six verbs used, which were 'lemmo', 'veemo', 'zoopo', 'neebo'. But the children's pointing trial performance was significantly worse than chance for only for two of these verbs ('lemmo' and 'zoopo'. The percentage of correct descriptions provided for 'pooko' and 'shanno' were greater than chance (although only 'shanno' was significant), which suggests that the children did understand the association between syntax and semantics for these clips and so this cannot be responsible for the children's poor performance at test with these verbs. Together these findings indicate that of the four SOV verbs that the children performed below chance on during the forced – choice comprehension test they only failed to understand the form to meaning mappings of two verbs ('lemmo' and 'zoopo'), which suggests that another factor may be responsible for the children's poor performance at test on the other two verbs ('shanno' and 'pooko'). However, it could be argued that the children had not yet developed the necessary metalinguistic skills to describe the clips correctly and so it is possible that they did in fact understand the meaning of the appearance films.

As of yet only the limited association between the children's post-test responses for the appearance verbs and their performance at test has been reviewed, however, there is a far stronger relationship between the mean percentage of SVO trials correct and the children's interpretations of these causative actions. The children demonstrated a clear understanding of the meaning of the SVO films by providing causative descriptions of the events significantly more often than chance for each verb, except 'neeb' (which showed a unique pattern of response that needs to be addressed separately). This comprehension of the majority of the causative clips may explain why the children performed significantly greater than chance during the transitive test trials for all the verbs, except 'veem' and 'neeb'. It also further implies that the children were biased towards linking syntax to semantics for the causative verbs, as proposed in previous research by Papafagou et al (2007) and Boyd et al (in press).

The children's poor performance on the appearance verbs in the forced – choice comprehension test is likely to result from their inadequate acquisition of the SOV training verbs as well as their poor understanding of the SOV test verbs. Evidence in support of this view can be found in the participant's descriptions of each of the appearance training scenes obtained during the translation post – test. Statistical analysis showed that the children gave correct responses significantly less often than chance for five of the eight SOV training films, which included all four instances of 'moopo' and 'koufo'. Above chance performance was only reported for one verb: 'faygo'. This indicates that the children were unable to acquire the majority of the appearance training verbs, which further explains why semantic analogies to new novel SOV verbs at test were so poor. The finding that the children failed to understand the meaning of 'moopo' is especially problematic since this verb was presented far more often than the others (four times as opposed to only once for the other verbs) with the intention that a single verb under high type frequency would aid the mapping of the novel construction to the meaning of appearance, as previously observed in Casenhiser and Goldberg (2005). It therefore seems likely that the children's inability to acquire the SOV training verbs is a major contributor to their poor performance on the appearance items at test.

The focus so far has been on demonstrating that because the children had insufficient understanding of the form to meaning mappings of the SOV test and training verbs they performed poorly on the appearance construction at test, however, it is also important to look at possible reasons why they were unable to comprehend the meaning of the verbs. For the children assigned to the replication condition the procedure was the same as that used in the original experiment conducted by Casenhiser and Goldberg (2005) except that in the present research all the children were given a translation post – test, but since this was carried out after the forced – choice comprehension test this addition to the procedure cannot be responsible for the children's failure to understand the form to meaning mappings of the SOV test verbs. Furthermore the individual verbs used at test and during the appearance training were identical to those used by Casenhiser and Goldberg (2005), so the characteristics of the words themselves (such as length or phonology) are unlikely to account for the children's inability to acquire their meanings. Clearly the procedure and verbs used are not the reason behind why the children in this study were unable to comprehend the meaning of the SOV verbs.

The most plausible explanation for why the children had such difficulty understanding the appearance verbs (test and training) is that the specific type of appearance meaning assigned to each different verb and the corresponding film clips designed to depict these events were not the same meanings and films used by Casenhiser and Goldberg (2005) and were not interpreted by the children as obviously demonstrating appearance, which meant analogies could not be made. The children showed particular problems with providing a description of appearance in the translation post – test for the training verb 'moopo' despite the fact that it was presented four times, compared to only once for the other SOV training verbs. The precise meaning given to 'moopo' was that of 'appearance out of thin air as if by magic'; with each of the clips first showing the scene without the animate subject and then with the subject placed on an inanimate object. When the researchers went back and reviewed the nature of the verbal descriptions given for 'moopo' it was found that the majority of they children had provided responses that focused on the final location of the subject, rather than its appearance into the scene. This suggests that either the meaning of 'appearance out of

thin air as if by magic' was too difficult for the children to understand or that the clips did not portray this unmistakably enough. Performance for 'moopo' could be improved by adding sound and lighting effects associated with magical appearance to the clips. The addition of such effects may also aid children's understanding of the other appearance films on which they performed significantly worse than chance. The only appearance items (test and training) for which the children's verbal descriptions were correct significantly more often than chance were 'shanno' and 'faygo'. 'Shanno' was assigned the meaning of 'popping out of a box', which is likely to be one of the most salient ways for something to appear for a child, since many children play with jack – in – a – boxes early on in life. Future studies in this area should also ask adults or older children to describe the film clips to ensure that participants can correctly interpret their meaning.

An inability to comprehend the specific meaning of the verb or the film depicting this event may also be responsible for the children's unusually poor performance on the transitive test verb 'neeb'. 'Neeb' was assigned the meaning of 'tipping a cup and pouring something into the sea to change its colour', which was a more complicated string of events than the actions described by many of the other novel causative verbs. This may have made it difficult for the children to identify the specific aspect of the scene to which 'neeb' referred to (Tomasello, 1995). Evidence in favour of this interpretation was obtained when the researchers reviewed the content of the verbal responses for 'neeb' given in the translation post – test and found that the majority of the children did not interpret the verb to mean changing the colour of the sea, but instead focused on the tipping up of the cup. This indicates that the children failed to understand that the intention of the agent (the shark) was to change the colour of the sea and not to simply tip the cup over, which may have altered the meaning they allocated to 'neeb'. Previous research by Poulin – Dubois and Forbes (2002) found that even at 27 months children's interpretations of a verb's meaning are influenced by what they perceive to be the intention of the subject. Furthermore the children were unable to provide the correct causal interpretation of 'neeb'; one possible explanation for this is that the concept of causality was linked to the object (the sea), whereas children prefer the subject (the shark) to be directly associated with the causation instead and so they based their responses on the tipping of the cup (Cohen & Oakes, 1993). Therefore it seems likely that the children's poor performance with the SVO test verb 'neeb' was related to their confusion over what the corresponding clip was demonstrating.

Throughout this discussion it has been consistently suggested that the children had difficulty understanding the form to meaning mappings of the appearance verbs. Recent research by Boyd and Goldberg (in press) propose that one way to help children acquire novel associations between syntax and semantics is to increase the lexical overlap between training and test constructions. These researchers found that children aged 4;6 to 6;8 performed significantly below chance at test with novel verbs of approach when the NP's in the sentence were different to those used in the training session (as in the present study), but significantly better than chance when the same subject and object were used across both phases of the study, as this highlights the similarity between the two events. Consequently, future studies should aim to maximise the amount of lexical overlap between sentences to speed up children's acquisition of form to meaning mappings.

The present research found that children aged 5;0 – 7;6 trained on a novel word order (SOV) associated with the meaning of appearance performed significantly worse

on new novel appearance verbs than novel causative verbs during a forced – choice comprehension test and so failed to replicate the findings of the second study by Casenhiser and Goldberg (2005). The most likely explanation for this finding is that the children in this study failed to understand the meaning of the appearance training and test verbs, because they did not perceive the accompanying film clips as showing appearance or comprehend the exact meaning of each individual appearance verb and so were unable to form semantic analogies across the exemplars.

References

Abbot –Smith, K., Lieven, E., & Tomasello, M. (2001). What preschool children do and do not do with ungrammatical word orders. *Cognitive Development, 16,* 679 – 692.

Akhtar, N. (1999). Acquiring basic word order: evidence for data driven learning of syntactic structure. *Journal of Child Language, 26 (2)*, 339 – 356.

Ambridge, B., Theakston, A., L., Lieven, E., V., M. & Tomasello, M. (2006). The distributed learning effect for children's acquisition of an abstract syntactic construction. *Cognitive Development, 21,* 174 – 193.

Ambridge, B., Pine, J., M., Rowland, C., F. & Young, C., R. (2008). The effect of verb semantic class and verb frequency (entrenchment) on children's and adults' graded judgments of argument structure overgeneralisation errors. *Cognition, 106,* 87 – 129.

Boyd, J., K. & Goldberg, A., E. (in press). Children's failure to generalise when exposed to the same input as adults and its benefits in language learning. *Lingua*.

Boyd, J., K., Gottschalk, E., A. & Goldberg, A., E. (in press). Linking rule acquisition in novel phrasal constructions. *Language Learning*.

Cameron – Faulkner, T., Lieven, E. & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science, 27,* 843 – 873.

Casenhiser, D., M. & Goldberg, A., E. (2005). Fast mapping between a phrasal form and meaning. *Developmental science, 8 (6)*, 500 – 508.

Chambers, K., Onishi, K., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition, 87,* B69 – B77.

Childers, J. & Tomasello, M. (2001). The role of pronouns in young children's acquisition of the English transitive construction. *Developmental Psychology, 37,* 739 – 748.

Childers, J. & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology, 38,* 967 – 978.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge; London: M.I.T. Press (1976 reprint).

Cohen, L., B. & Oakes, L., M. (1993). How infants perceive a simple causal event. *Developmental Psychology, 29 (3),* 421 – 433.

Dodson, K. & Tomasello, M. (1998). Acquiring the transitive construction in English: The role of animacy and pronouns. *Journal of Child Language, 25*, 555 – 574.

Fernandes, K., Marcus, G., F., DiNubila, J., A., & Vouloumanos, A. (2006). From semantics to syntax and back again: Argument structure in the third year of life, *Cognition, 100 (2),* B10 – B20.

Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological Science, 17 (8),* 684 – 691.

Gentner, D. & Namy, L., L. (2006). Analogical processes in language learning. *Current Directions in Psychological Science, 15 (6),* 297 – 301.

Goldberg, A., E., Casenhiser, D., M. & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics, 15 (3),* 289 – 316.

Goldberg, A., E. (2006). *Constructions at work: the nature of generalisation in language.* New York; Oxford: Oxford University Press.

Gomez, R., L. & Gerken, L. (1999). Artificial grammar learning by 1- year – olds leads to specific and abstract knowledge. *Cognition, 70 (2),* 109 – 135.

Ninio, A. (1999). Pathbreaking verbs in syntactic development and the question of prototypical transitivity. *Journal of Child Language, 26,* 619 – 653.

Ninio, A. (2005). Testing the role of semantic similarity in syntactic development. *Journal of Child Language, 32,* 35 – 61.

Papafragou, A., Cassidy, K. & Gleitman, L. (2007). When we think about thinking: The acquisition of belief verbs. *Cognition, 105,* 125 – 165.

Pinker, S. (1984). *Language learnability and language development*. Cambridge; London: Harvard University Press.

Poulin – Dubois, D. & Forbes, J., N. (2002). Toddlers' attention to intentions – in – action in learning novel action words. *Developmental Psychology, 31 (1),* 104 – 114.

Ramscar, M. (2002). The role of meaning in inflection: why the past tense does not require a rule. *Cognitive Psychology, 45 (1),* 45 – 94.

Tomasello, M. (1995). *Beyond names for things: Young children's acquisition of* verbs. Hillsdale, NJ: L. Erlbaum.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition, 74*, 209 – 253.

Tomasello, M. (2003). *Constructing a language: a usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Van der Lely, H., K., J. (1994). Canonical linking rules: forward versus reverse linking in normally developing and specifically language impaired children. *Cognition, 51 (1),* 29 – 72.

Appendices

The appendices to this report can be viewed in the folder 'Supplementary Files' located in the Reading Tools menu list.